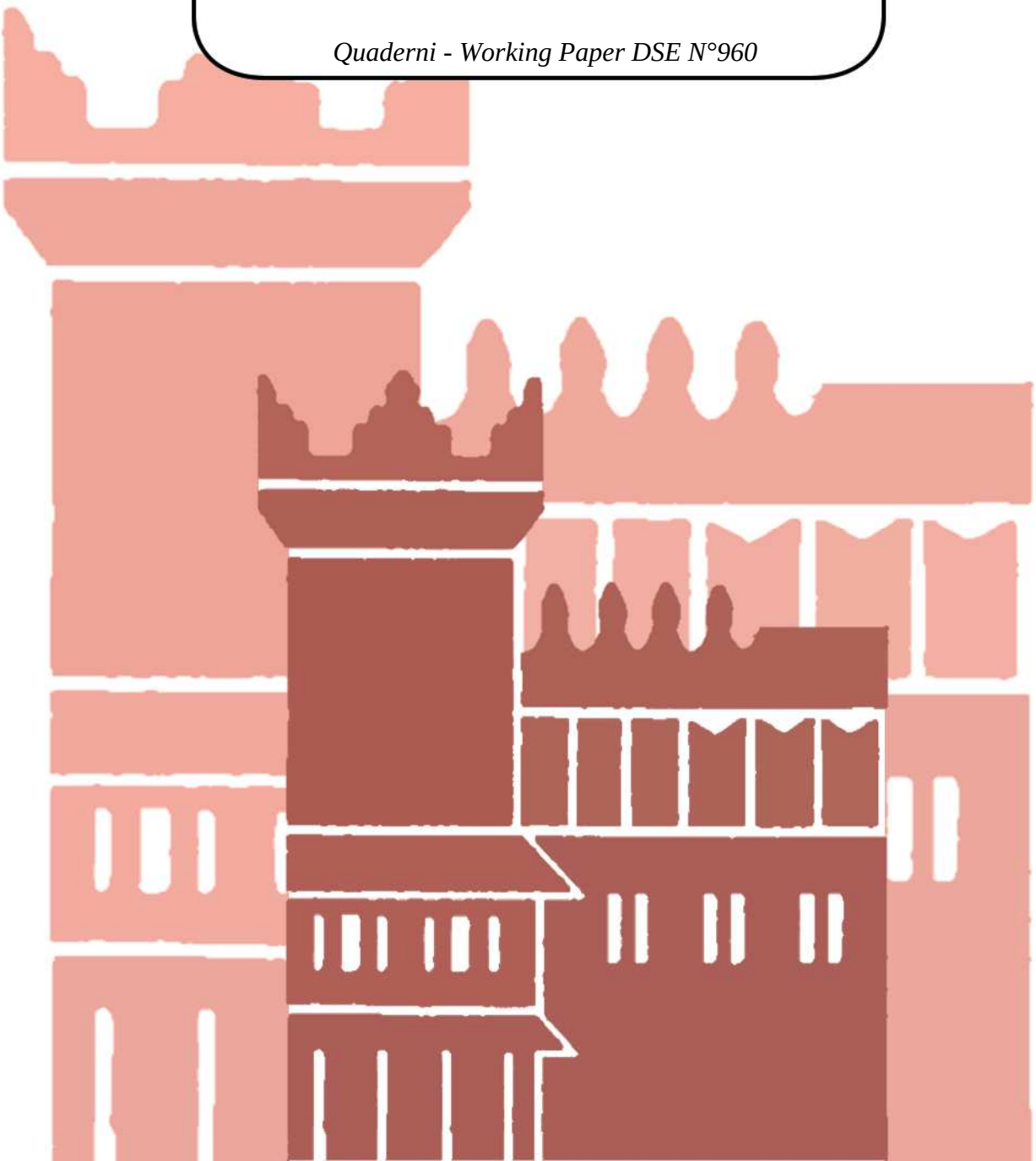
ISSN 2282-6483

Alma Mater Studiorum - Università di Bologna
DEPARTMENT OF ECONOMICS

pca2: implementing a strategy to reduce the instrument count in panel GMM

Maria Elena Bontempi
Irene Mammi

Quaderni - Working Paper DSE N°960



pca2: implementing a strategy to reduce the instrument count in panel GMM

Maria Elena Bontempi*
Irene Mammi†

August 27, 2014

Abstract

The problem of instrument proliferation and its consequences (overfitting of the endogenous explanatory variables, biased IV and GMM estimators, weakening of the power of the overidentification tests) are well known. This paper introduces a statistical method to reduce the instrument count. The principal component analysis (PCA) is applied on the instrument matrix, and the PCA scores are used as instruments for the panel generalized method-of-moments (GMM) estimation. This strategy is implemented through the new command pca2.

Keywords: pca2, proliferation of instruments, principal component analysis, panel data, generalized method of moments.

*Dep. of Economics, University of Bologna. Email: mariaelena.bontempi@unibo.it
†Dep. of Economics, University of Bologna. Email: irene.mammi@unibo.it

1 Introduction

The generalized method-of-moments (GMM) estimator, in the Arellano and Bond [1991], Arellano and Bover [1995] and Blundell and Bond [1998] formulations, has gained a leading role among the dynamic panel data (DPD) estimators, mainly due to its flexibility and to the few assumptions about the data generating process it requires. In addition, the availability of lags of the endogenous variables provides many instrumental variables (IVs) directly exploitable for GMM estimation.

However, the estimation of DPD models by GMM with many instruments has its own drawbacks. Already in the seminal work of Sargan [1958] it was stressed that, in the context of IV estimation, the marginal improvements from an increase in the number of instruments beyond three are generally small whereas they can negatively affect the consistency of the estimates and the reliability of specification tests. Since then, the potential distortions in parameter estimates when the instrument count gets larger have been further extensively investigated in the literature (Kiviet [1995], Andersen and Sorensen [1996], Ziliak [1997] among others).

In particular, instrument proliferation is intrinsic in GMM estimation of DPD models when all the lags of the endogenous explanatory variables are exploited, as the number of moment conditions increases with T and with the dimension of the vector of endogenous regressors. While, in principle, the availability of a wider set of conditions should improve efficiency (Dagenais and Dagenais [1997]), the bias due to overfitting is quite severe as the number of moment conditions expands, outweighting the gains in efficiency (Bekker [1994], Newey and Smith [2004], Ziliak [1997]). Such trade-off between bias and efficiency is exacerbated by the weak instruments problem (Bound et al.[1995], Staiger and Stock [1997]) and by the correlation between the sample moments and the estimated optimal weighting matrix, as sampling errors are magnified in the weighting matrix (Altonji and Segal [1994]). Poor estimates of the variance/covariance matrix of the moments lower the power of the specification tests such as the Sargan/Hansen test for overidentifying restrictions, that suffers from a severe under-rejection problem (Sargan [1958], Andersen and Sorensen [1996], Bowsher [2002]).

Overall, such evidence supports the importance of properly addressing instruments proliferation, although this problem is often overlooked in empirical analyses; indeed only seldom empirical strategies such as lag truncation and collapse (Roodman [2009a]) are used to reduce the number of moment conditions. This reduction is especially required when the number of IVs is small with respect to the cross-sectional dimension of the panel (Alvarez and Arellano [2003]).

Our aim in this paper is to tackle the issue of instrument proliferation by providing a statistically grounded and directly implementable procedure

that reduces the instrument count. In line with Doran and Schmidt (2005) who propose an eigenvalue-eigenvector decomposition of the GMM weighting matrix in order to reduce its dimension, we advocate the use of principal components analysis (PCA) of the instruments matrix as a way to shrink the available instruments into a set of linear combinations of the original variables (the scores of the PCA). The weights used in such orthogonal combinations follow from the main features of the data and reflect the contribution of each variable to the total observed variability.

We label this strategy "principal components instrumental variables reduction" (*PCIVR*). *PCIVR* comes out as a complementary tool with respect to lag truncation and collapse, which impose *a priori* restrictions not tailored on the data. Instead the approach we propose here provides a flexible statistical rule for the selection of non redundant instruments that adjusts to the empirical problem at hand and reflects the specific features of the data[1].

Our procedure extends the set of tools available to the researcher to reduce the instrument count, as it embraces a different perspective with respect to the existing strategies. It should be recalled that lag truncation assumes that the relevant information is only conveyed by the most recent (usually one or two) available lags of the endogenous variables, while the collapsing of the instruments matrix assumes specific dynamics in the data. As such assumptions can not be tested a priori, in order to identify potential critical aspects related to the issue at hand, we believe that it is highly advisable to compare the GMM estimates obtained with lag truncation and collapsing with those provided by *PCIVR*, where the information from the whole instruments set is exploited in order the select the lags that contribute most to the total variability. Given that none of the count reduction strategies mentioned above can be shown to be superior to the others in every situation, further robustness analysis is recommended in this context. This robustness checks can now be easily done through the new `pca2` command, that directly implements *PCIVR* by means of an .ado file.

Thanks to its flexibility, the command `pca2` adds useful features to the Stata command `pca`. It is straightforwardly applicable in Stata to any type of dataset (cross-section, time series and panel), and automates, through specific options, alternative ways to extract the principal components and to select those to be retained for the computation of the PCA.

The rest of the paper is organised as follows. Section 2 summarises the main methodological underpinnings of the strategy we present: section 2.1 reviews the GMM estimation of DPD models and section 2.2 describes the extraction of principal components from a matrix of instruments. Section 3 details the syntax of the `pca2` and its options by also providing some empirical

[1] A first sketch of a PCA-based reduction of GMM IVs can be found in Mehrhoff [2009], while the `pca` option in the user-written `xtabond2` (Roodman [2009b]) command provides a first application within Stata.

examples. Section 4 carries out a guided example of robustness analysis in the context of published research on the determinants of the discretionary fiscal policy in the Euro area Countries.

2 The methodological framework

2.1 GMM estimation of DPD models

Consider the general two-way error component DPD model:

$$y_{it} = \alpha y_{it-1} + \boldsymbol{\beta}' \mathbf{x}_{i,t} + \phi_t + v_{it} \ , \ v_{it} = \eta_i + \varepsilon_{it}, \tag{1}$$

where $i = 1, .., N$, $t = 1, .., T$, $\mathbf{x}$ is a m-dimensional vector of potentially endogenous or predetermined regressors, the ϕ_t are the time effects, the η_i are the individual effects and ε_{it} is a zero-mean idiosyncratic error, allowed to be heteroskedastic but not serially correlated. The standard assumptions are: $E[\eta_i]$=$E[\varepsilon_{it}]$=$E[\eta_i \varepsilon_{it}]$=0 and predetermined initial conditions $E[y_{i1}\varepsilon_{it}]$=0.

The Arellano-Bond and Arellano-Bover/Blundell-Bond estimators are linear GMM estimators for the model in (1) in first differences (DIF GMM) or in levels (LEV GMM) or both (SYS GMM); the instrument matrix $\mathbf{Z}$ includes the lagged values of the endogenous variables. The columns of $\mathbf{Z}$ correspond respectively to two different sets of meaningful moment conditions.

The Arellano-Bond DIF GMM estimator exploits the following moment conditions for the equation (1) in first differences:

$$E[(\mathbf{Z_i}^{\text{dif}})' \Delta v_i] = E[(\mathbf{Z_{i,t-l}}^{\text{dif}})' \Delta v_{it}] = 0 \text{ for } t \geq 3, l \geq 2. \tag{2}$$

where l denotes the lag depth.

The Blundell-Bond SYS GMM estimator also exploits the additional non-redundant orthogonality conditions for the equation (1) in levels:[2]

$$E[(\mathbf{Z_i}^{\text{lev}})' v_i] = E[(\mathbf{Z_{is}}^{\text{lev}})' v_{iT}] = 0 \text{ for } s = 2, ..., T-1. \tag{3}$$

Since DPD GMM uses lags of the explanatory variables as IVs, "the phenomenon of moment condition proliferation is far from being a theoretical construct and arises in a natural way in many empirical econometric settings" (Han and Phillips [2006, p. 149]). The dimension of the GMM-type instrument matrix grows as the number of time periods and endogenous regressors expands.

[2]The LEV GMM estimation considers, for each endogenous variable, time period and lag distance, all the available lags of the first differences as instrument for the equation in levels because they are non redundant.

2.2 Extracting principal components from the matrix of instruments

The adoption of PCA or factor analysis to extract a small number of factors from a large set of variables has become popular in macroeconometrics, the forecasting being the main field of application. Stock and Watson [2002] prove consistency of the factors as the number of original variables gets sufficiently large, so that the principal components are estimated precisely enough to be used as data instead of the original variables in subsequent regressions. Kloek and Mennes [1960] and Amemiya [1966] first propose the use of principal components in the IV estimation. Important recent contributions, among the others, are Kapetanios and Marcellino [2010], Groen and Kapetanios [2009] and Bai and Ng [2010][3].

The issue of instruments proliferation can be addressed by extracting the principal components from the instrument matrix $\mathbf{Z}$. The aim of PCIVR is to re-express the information conveyed by highly correlated variables in terms of a set of optimal orthogonal linear combinations of the original variables and then to retain a smaller number of them.

In detail, defined $\mathbf{Z}$ as the general p-columns[4] GMM-style instrument matrix, we extract p eigenvalues $\lambda_1, \lambda_2, ..., \lambda_p \geq 0$ from the correlation or covariance matrix of $\mathbf{Z}$, ordered from the largest to the smallest, and derive the corresponding eigenvectors (principal components) $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_p$. Our new instruments will be the scores from PCA that are defined as:

$$\mathbf{s}_k = \mathbf{Z}\mathbf{u}_k \text{ for } k = 1, 2, ..., p. \tag{4}$$

If we write $\mathbf{Z} = [\mathbf{z}_1 \quad ... \quad \mathbf{z}_j \quad ... \quad \mathbf{z}_p]$ with $\mathbf{z}_j$ being the j^{th} column of the instrument matrix, the score $\mathbf{s}_k$ corresponding to the k^{th} component can be rewritten as:

$$\mathbf{s}_k = u_{k1}\mathbf{z}_1 + ... + u_{kj}\mathbf{z}_j + ... + u_{kp}\mathbf{z}_p \tag{5}$$

where u_{kj} is the j^{th} element of the principal component $\mathbf{u}_k$. Defined the matrix of PCA loadings as $\mathbf{V} = [\mathbf{u}_1 \quad ...\mathbf{u}_k \quad ... \quad \mathbf{u}_p]$ and the matrix of PCA scores as $\mathbf{S}$, we have that $\mathbf{S} = \mathbf{Z}\mathbf{V}$. Instead of the moment conditions in (2), we will therefore exploit the following restrictions in GMM DIF:

$$E[(\mathbf{S}^{\text{dif}})'\Delta\boldsymbol{v}] = E[(\mathbf{Z}^{\text{dif}}\mathbf{V})'\Delta\boldsymbol{v}] = 0. \tag{6}$$

Similarly, in the GMM SYS we will also exploit the additional orthogonality conditions

$$E[(\mathbf{S}^{\text{lev}})'\boldsymbol{v}] = E[(\mathbf{Z}^{\text{lev}}\mathbf{V})'\boldsymbol{v}] = 0. \tag{7}$$

[3] A review of the literature on Factor-IV and Factor-GMM estimations is in the introduction of Kapetanios and Marcellino [2010].

[4] The matrix $\mathbf{Z}$ has p columns that can either be the lags -in levels, first-differences or both- of a single variable taken separately from the others or they can be the lags -in levels, first-differences or both- of more variables considered together. All these possibilities are directly implementable using the `pca2` command.

Since the aim of the PCIVR is the reduction of the dimension of the instruments matrix, a criterion to select the scores to be retained has to be adopted. The idea is retaining only $(m+1) \leq q < p$ principal components, where m is the number of endogenous regressors other than the lagged dependent variable; as a consequence, only the q corresponding score vectors will form the new transformed instrument matrix in both (6) and (7).

One possibility is to retain the q principal components corresponding to eigenvalues above the average of the eigenvalues (Average criterion); alternatively, one may keep those accounting for a given percentage of the variance of the data, generally 70% to 90% (Variability criterion).

The number of moment restrictions resulting from the *PCIVR* depends on the nature of the data at hand. If $q < (m+1)$, the equation of interest is not identified. This can happen for instance when the variables are highly persistent (near unit root processes): in this case, the PCA is driven by spurious trends and too few principal components are retained.

3 pca2 command: syntax

The user-written command `pca2` implements the *PCIVR* procedure presented above: in a unique step, it extracts the principal components from the variables in the *varlist* according to the preferences specified through its options; then it computes the scores corresponding to the principal components retained on the basis of the selection criterion chosen by the researcher. These scores can be used in any IV/GMM estimation command in Stata in place of the original IVs.

The extraction of principal components through the `pca2` command exploits the Stata `pca` command. Its innovative feature consists of augmenting the `pca2` command with specific options for the creation of GMM-style IVs, for the selection of principal components and for the computation of the scores.

The syntax of `pca2` is

`pca2` *varlist* `[if` *exp*`] [in` *range*`] [, nt(`*timevar* | *panelvar timevar*`)`
`variance(#) avg covariance prefix(`*string*`) see gmmliv(# | # #)`
`gmmdiv(# | # #) lagsl(`*varistl*`,` *ll*`(# | # #)) lagsd(`*varistd*`,` *ll*`(# | # #))`
`togvar togld retain ]`.

Time-series and panel data must be `tsset` before using `pca2`. See `[TS] tsset` for more information. `pca2` does not allow time-series operators in the *varlist*; in order to use lags of the variables in the *varlist*, they need to be generated using Stata time-series operators before applying the `pca2` command (see `help tsvarlist`).

3.1 pca2: options

`nt(`*timevar* | *panelvar timevar*`)` is required in time-series and panel data in order to create GMM-style instruments and to apply PCA on them. If this option is omitted, the dataset is treated as a cross-section and all the observations would be pooled.

`variance(#)` allows to apply the "variance criterion" (default criterion) i.e. only those principal components that account for at least the chosen percentage of the variability in the original data are retained for the computation of the scores. The number defining the percentage must be an integer greater than 0 and lower or equal to 100. The default is # = 90.

`avg` selects the principal components to be kept for score computation according to the "average criterion", i.e. only those eigenvectors whose corresponding eigenvalues are above the average of the eigenvalues are retained. Note that when the options `avg` is chosen, `pca2` also computes the scores according to the default 90% "variance criterion" and saves both of them in the dataset: the scores obtained according to the two criteria can thus be compared.

`covariance` performs PCA of the covariance matrix; default is to perform PCA on the correlation matrix (see `help pca`).

`prefix(`*string*`)` specifies the prefix for the name of the scores generated by the `pca2` command corresponding to the retained principal components. If you write e.g. `prefix(`*sys*`)` you will obtain _sysvarscore* and _sysavgscore*. This option is particularly useful when the `pca2` command is repeated many times on the same dataset in order to create different scores from different instrument sets, eventually also according to different criteria. The default prefix is _BM which retains the scores with labels such as _BMvarscore* and _BMavgscore*.

`see` asks Stata to display the outcome of the PCA.

`gmmliv(# | # #)` generates the GMM-style instruments in levels (for the equations in first-differences) for all the variables included in the `pca2` *varlist*. If only one argument is specified, e.g. `gmmliv(`k`)`, all the available lags from $t-k$ back to the initial observation for each variable in the varlist of the `pca2` command are used. If two arguments are specified, e.g. `gmmliv(`$k1$ $k2$`)` with $k1 \leq k2$, the lags from $t-k1$ to $t-k2$ are considered. The PCA is done on all the specified GMM-style lags in levels of each variable taken separately. If the option `togvar` (full description below) is also added, the PCA is performed on all the generated GMM-style lags in levels of all the variables in the *varlist* considered together. With this option the lag structure is the same for each variable.

`gmmdiv(# | # #)` generates the GMM-style instruments in first-differences (for the equations in levels) for all the variables included in the `pca2` *varlist*. If only one argument is specified, e.g. `gmmdiv(`k`)`, all the available lags from $t-k$ back to the initial observation for each variable in the *varlist* of the `pca2`

command are used. If two arguments are specified, e.g. `gmmdiv(`$k1$ $k2$`)` with $k1 \leq k2$, the lags from $t - k1$ to $t - k2$ are considered. The PCA is done on all the specified GMM-style lags in first-differences of each variable taken separately. If the option `togvar` (full description below) is also added, the PCA is performed on all the generated GMM-style lags in first-differences of all the variables in the *varlist* considered together. With this option the lag structure is the same for each variable.

`lagsl(`*varlistl*`,` *ll*`(# | # #))` generates the GMM-style instruments in levels for a specific *varlistl*. It is a more flexible alternative to the `gmmliv()` option, as it allows for a different lag structure of each variable. The option `lagsl()` may be used more than once: different lag structures may thus be defined for the variables in each *varlistl*. The sub-option *ll* specifies the lag structure of the variables in each *varlistl*: if only one argument is specified, e.g. *ll*`(`k`)`, all the available lags from $t - k$ back to the initial observation for each variable in the *varlistl* are used. If two arguments are specified, e.g. *ll*`(`$k1$ $k2$`)` with $k1 \leq k2$, the lags from $t - k1$ to $t - k2$ are considered. The PCA is done on all the specified GMM-style lags in levels of each variable taken separately. If the option `togvar` (full description below) is also added, the PCA is performed on all the generated GMM-style lags in levels of all the variables in the *varlistl* considered together. Such option can not be used with the option `gmml()`, while it is allowed together with either the option `lagsd()` or with the option `gmmd()`. When used alone or with the option `lagsd()` the number of variables in both *varlistl* and *varlistd* must be at least equal to the number of the variables in the *varlist* of the `pca2` command. Only when associated with the option `gmmd()`, the `lagsl()` option can have fewer variables than those included in the *varlist* of the `pca2` command.

`lagsd(`*varlistd*`,` *ll*`(# | # #))` generates the GMM-style instruments in first-differences for a specific *varlistd*. It is a more flexible alternative to the `gmmdiv()` option, as it allows for a different lag structure of each variable. The option `lagsd()` may be used more than once: different lag structures may thus be defined for the variables in each *varlistd*. The sub-option *ll* specifies the lag structure of the variables in each *varlistd*: if only one argument is specified, e.g. *ll*`(`k`)`, all the available lags from $t - k$ back to the initial observation for each variable in the *varlistd* are used. If two arguments are specified, e.g. *ll*`(`$k1$ $k2$`)` with $k1 \leq k2$, the lags from $t - k1$ to $t - k2$ are considered. The PCA is done on all the specified GMM-style lags in first-differences of each variable taken separately. If the option `togvar` (full description below) is also added, the PCA is performed on all the generated GMM-style lags in first-differences of all the variables in the *varlistd* considered together. Such option can not be used with the option `gmmd()`, while it is allowed together with either the option `lagsl()` or with the option `gmml()`. When used alone or with the option `lagsl()` the number of variables in both *varlistl* and *varlistd* must be at least equal to the number of the variables in the *varlist* of the `pca2` command. Only

when associated with the option `gmml()`, the `lagsd()` option can have fewer variables than those included in the *varlist* of the `pca2` command.

`togvar` specifies that the PCA is performed on the matrix that includes all the variables in the *varlist* and not on each variable separately. E.g. the syntax `pca2 x z, togvar` implies that the PCA is performed jointly on the variables x and z. This option needs to be specified in order to apply the PCA to GMM-style lags of more than one variable taken together instead of the lags of each variable taken separately. For example, `pca2 x z, gmml(2) togvar` implies that the principal components are extracted from the matrix that includes all the available lags in levels from $t-2$ backward of the variables x and z.

`togld` specifies that, once that instruments in levels and first-differences are generated, the PCA is applied to the matrix that includes all these instruments together for each variable in the *varlist* of the `pca2`. If the option `togld` is used together with the option `togvar`, the principal components are extracted from the matrix that includes all the lags in first-differences and in levels of all the variables in the *varlist*.

`retain` adds the generated GMM-style instrumental variables as new variables in the dataset. These IVs are named *_GMMLvarname*period*lag*; for example *_GMMLn*1978*L*2 stands for the $t-2$ observation in levels for the variable n in the year $t = 1978$.

For a description of the full set of options accepted by the `pca2` command, type `help pca2`.

3.2 The use of the `pca2` command: an example

We illustrate the command `pca2` through an empirical example based on the `abdata.dta` dataset used in Arellano and Bond [1991] and Blundell and Bond [1998].

We estimate the Blundell and Bond [1998] model, a simple autoregressive distributed lags model of labor demand:

$$n_{it} = \alpha n_{it-1} + \beta_0 w_{it} + \beta_1 w_{it-1} + \gamma_0 k_{it} + \gamma_1 k_{it-1} + \eta_i + \phi_t + \nu_{it} \qquad (8)$$

where n_{it}, w_{it} and k_{it} are the log of employment, the log of the real product wage and the log of the capital stock in firm i in year t, respectively. The sample is an unbalanced panel of 140 UK listed manufacturing companies with between 7 and 9 annual observations over the period 1976-1984.

First, we replicate the original DIFF GMM results in column 3 of Table 4 in Blundell and Bond [1998]; then, we estimate the same model by DIFF GMM estimates exploiting the set of IVs resulting from the PCA.

To do so, we run the command:

```
pca2 n w k, nt(id year) gmml(2) retain avg.
```

This syntax generates the GMM-style instruments in levels for each of the n, w and k variables. These variables are labeled $\_GMML^*$ and will be used as instruments for equation (8) in first-differences. In this case, the option `gmml(2)` specifies the same lag structure for all the variables and the IVs are generated from lag $t-2$ up to the last lag available. By specifying the option `retain`, the $\_GMML^*$ instruments are added as new variables in the dataset.

Then, the principal components are separately extracted for each variable from its own lags; next, the principal components are retained according to both selection criteria (i.e. the default "variance criterion" and the "average criterion"), as the option `avg` is used; the corresponding scores are then saved in the dataset as new variables labelled $\_BM\_var^*$ and $\_BM\_avg^*$, where *var* and *avg* refers to the selection criterion.

As shown right below, the output of the `pca2` command reports information about the lag structure of the GMM-style IVs and summary statistics for the extraction of the principal components.

```
. use http://fmwww.bc.edu/ec-p/data/macro/abdata.dta, clear

. xtset id year

. quietly tab  year, gen(tauyear)

. pca2 n w k, nt(id year) gmml(2) retain avg
General description of the dataset
       panel variable:  id (unbalanced)
        time variable:  year, 1976 to 1984
                delta:  1 unit
The prefix is: _BM_
You are creating GMM-style IVs in levels for a panel
_____ variable: n _____
Lag selection in GMML(): from t-2 to the last available lag
_____ variable: w _____
Lag selection in GMML(): from t-2 to the last available lag
_____ variable: k _____
Lag selection in GMML(): from t-2 to the last available lag

_____ PCA LEV VAR BY VAR: n
You are applying PCA to GMM-style LEV lags of one or more than one variable,
keeping the variables separated with the same lags structure
 _________ Some information about PCA of IV in levels for n __________
Trace of the matrix:                                                  28
By default percentage of selected variability to be explained:        90%
Percentage of variance explained by the variability criterion:        92.943733%
Number of retained scores according to the variability criterion:     8
Percentage of variance explained by the average criterion:            86.399506%
Number of retained scores according to the average criterion:         6

_____ PCA LEV VAR BY VAR: w
You are applying PCA to GMM-style LEV lags of one or more than one variable,
keeping the variables separated with the same lags structure
```

```
 _________ Some information about PCA of IV in levels for w __________
Trace of the matrix:                                                    28
Percentage of variance explained by the variability criterion:          90.305677%
Number of retained scores according to the variability criterion:       7
Percentage of variance explained by the average criterion:              87.588082%
Number of retained scores according to the average criterion:           6

_____ PCA LEV VAR BY VAR: k
You are applying PCA to GMM-style LEV lags of one or more than one variable,
keeping the variables separated with the same lags structure
 _________ Some information about PCA of IV in levels for k __________
Trace of the matrix:                                                    28
Percentage of variance explained by the variability criterion:          90.223503%
Number of retained scores according to the variability criterion:       7
Percentage of variance explained by the average criterion:              86.652737%
Number of retained scores according to the average criterion:           6
```

In order to get the original DIFF GMM estimates for the model in (8), we can use the user written `xtabond2` command (see Roodman [2009b]) with its native syntax:

```
quietly xtabond2 n l.n l(0/1).(w k) tauyear3-tauyear9, //
iv(tauyear3-tauyear9, eq(diff)) gmm(n, lag(2 .) eq(diff)) //
gmm(w, lag(2 .) eq(diff)) gmm(k, lag(2 .) eq(diff)) //
h(2) nolev robust nod.
```

However, in order to illustrate how to exploit the instrumental variables obtained through the `pca2` command, in this case the $\_GMML^*$ IVs just added to the dataset, we can reproduce the same estimates by typing:

```
xtabond2 n l.n l(0/1).(w k) tauyear3-tauyear9, //
iv(tauyear3-tauyear9, eq(diff)) iv(_GMML_n_*, eq(diff) pass) //
iv(_GMML_w_*, eq(diff) pass) iv(_GMML_k_*, eq(diff) pass)  //
h(2) nolev robust nod
```

where the new variables are used as traditional IVs through the option `ivstyle`. The output of the two commands is exactly the same. The output of the second command above is reported here.

```
Dynamic panel-data estimation, one-step difference GMM
------------------------------------------------------------------------------
Group variable: id                              Number of obs      =       751
Time variable : year                            Number of groups   =       140
Number of instruments = 91                      Obs per group: min =         5
Wald chi2(12) =   1163.33                                      avg =      5.36
Prob > chi2   =     0.000                                      max =         7
------------------------------------------------------------------------------
             |               Robust
           n |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
```

```
-------------+----------------------------------------------------------------
           n |
         L1. |   .7074701   .0841788     8.40   0.000     .5424827    .8724576
             |
           w |
         --. |  -.7087965    .117102    -6.05   0.000    -.9383122   -.4792809
         L1. |   .5000149   .1113282     4.49   0.000     .2818157    .7182141
             |
           k |
         --. |   .4659776    .101044     4.61   0.000      .267935    .6640203
         L1. |  -.2151309   .0858525    -2.51   0.012    -.3833987   -.0468631
             |
    tauyear3 |   .0057636   .0166077     0.35   0.729    -.0267868     .038314
    tauyear4 |   .0136366   .0193748     0.70   0.482    -.0243374    .0516106
    tauyear5 |  -.0071557   .0213479    -0.34   0.737    -.0489969    .0346855
    tauyear6 |  -.0340692   .0264327    -1.29   0.197    -.0858763    .0177379
    tauyear7 |  -.0059175   .0272325    -0.22   0.828    -.0592922    .0474573
    tauyear8 |   .0187213   .0288529     0.65   0.516    -.0378294     .075272
    tauyear9 |   .0352279   .0331578     1.06   0.288    -.0297603    .1002161
------------------------------------------------------------------------------
Instruments for first differences equation
  Standard
    _GMML_k_1978L2 _GMML_k_1979L2 _GMML_k_1979L3 _GMML_k_1980L2 _GMML_k_1980L3

    (output omitted)

    _GMML_n_1984L6 _GMML_n_1984L7 _GMML_n_1984L8
    D.(tauyear3 tauyear4 tauyear5 tauyear6 tauyear7 tauyear8 tauyear9)
------------------------------------------------------------------------------
Arellano-Bond test for AR(1) in first differences: z =  -5.60  Pr > z =  0.000
Arellano-Bond test for AR(2) in first differences: z =  -0.14  Pr > z =  0.891
------------------------------------------------------------------------------
Sargan test of overid. restrictions: chi2(79)   = 125.19  Prob > chi2 =  0.001
  (Not robust, but not weakened by many instruments.)
Hansen test of overid. restrictions: chi2(79)   =  88.80  Prob > chi2 =  0.211
  (Robust, but weakened by many instruments.)
```

We now estimate the model in equation (8) by DIFF GMM on the set of instruments that results from *PCIVR*. The `pca2` run above saves the PCA scores *_BM_varscoreDIF** and *_BM_avgscoreDIF** as new variables in the dataset. Therefore, we can get the estimates on the new set of instruments by using, for example, the variables *_BM_var** in `xtabond2` as new instruments instead of the standard ones as follows:

```
xtabond2 n l.n l(0/1).(w k) tauyear3-tauyear9, //
iv(tauyear3-tauyear9, eq(diff)) iv(_BM_var*n*, eq(diff) pass) //
iv(_BM_var*w*, eq(diff) pass) iv(_BM_var*k*, eq(diff) pass) //
h(2) nolev robust nod.
```

The estimation results are presented below.

```
Dynamic panel-data estimation, one-step difference GMM
------------------------------------------------------------------------------
Group variable: id                              Number of obs      =       751
Time variable : year                            Number of groups   =       140
Number of instruments = 29                      Obs per group: min =         5
Wald chi2(12) =   1146.02                                      avg =      5.36
Prob > chi2   =     0.000                                      max =         7
------------------------------------------------------------------------------
             |               Robust
           n |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           n |
         L1. |   .8021886   .1255146     6.39   0.000     .5561845    1.048193
             |
           w |
         --. |  -.8621674   .2094745    -4.12   0.000     -1.27273   -.4516048
         L1. |   .2224614   .2941419     0.76   0.449    -.3540461     .798969
             |
           k |
         --. |   .5783907   .2253891     2.57   0.010     .1366362    1.020145
         L1. |  -.4108413   .1947894    -2.11   0.035    -.7926216    -.029061
             |
    tauyear3 |  -.0202252   .0272124    -0.74   0.457    -.0735604      .03311
    tauyear4 |  -.0114123   .0355594    -0.32   0.748    -.0811074    .0582829
    tauyear5 |  -.0209936   .0374262    -0.56   0.575    -.0943475    .0523603
    tauyear6 |   -.034543    .049461    -0.70   0.485    -.1314848    .0623988
    tauyear7 |   .0148526   .0524715     0.28   0.777    -.0879897    .1176949
    tauyear8 |   .0556274   .0447092     1.24   0.213     -.032001    .1432558
    tauyear9 |   .0688565   .0555122     1.24   0.215    -.0399454    .1776584
------------------------------------------------------------------------------
Instruments for first differences equation
  Standard
    _BM_varscoreLEVkN1 _BM_varscoreLEVkN2 _BM_varscoreLEVkN3

    (output omitted)

    _BM_varscoreLEVnN7 _BM_varscoreLEVnN8
    D.(tauyear3 tauyear4 tauyear5 tauyear6 tauyear7 tauyear8 tauyear9)
------------------------------------------------------------------------------
Arellano-Bond test for AR(1) in first differences: z =  -3.41  Pr > z =  0.001
Arellano-Bond test for AR(2) in first differences: z =  -0.61  Pr > z =  0.544
------------------------------------------------------------------------------
Sargan test of overid. restrictions: chi2(17)   =  32.49  Prob > chi2 =  0.013
  (Not robust, but not weakened by many instruments.)
Hansen test of overid. restrictions: chi2(17)   =  23.43  Prob > chi2 =  0.136
  (Robust, but weakened by many instruments.)
```

As the aim of the *PCIVR* is the reduction in the instrument count, as expected the Hansen test has 79 degrees of freedom in the standard DIF GMM estimates, while they fall to 17 when the score relative to the principal components are extracted according to the variance criterion.

So far we have focused on the syntax for the use of GMM-style instruments

and PCA scores in the DIFF GMM estimation. In addition to that, we can also use the `pca2` to create IVs and PCA scores to be used in SYS GMM[5]; we can thus replicate the results in column 4 of Table 4 in Blundell and Bond [1998] and get SYS GMM estimates with the PCA scores as instruments.

The syntax

```
pca2 n w k, nt(id year) gmml(2) gmmd(1 1) retain avg
```

creates both the instruments in levels from $t-2$ up to the last lag available and first-differences for the first available lag. The IVs in levels, i.e. the *_GMML** variables, and the instruments in first-differences, i.e. the *_GMMD** variables, are included in the dataset as new variables. The PCA is run on the instruments in first-differences and on the instruments in levels for each variable separately; the scores relative to the retained principal components (*_BM_varscoreDIF** and *_BM_avgscoreDIF**, *_BM_varscoreLEV** and *_BM_avgscoreLEV**) are also added to the dataset.

```
(output omitted)
```

Following the same line of reasoning, we can get the standard SYS GMM estimates by using the `xtabond2` with the *_GMMD** and *_GMML** variables as instruments through the command:

```
xtabond2 n l.n l(0/1).(w k) tauyear3-tauyear9, ///
iv(tauyear3-tauyear9, eq(both))  iv(_GMML_n_*, eq(diff) pass) ///
iv(_GMML_w_*, eq(diff) pass) iv(_GMML_k_*, eq(diff) pass) ///
iv(_GMMD_n_*L1, eq(lev)) iv(_GMMD_w_*L1, eq(lev)) ///
iv(_GMMD_k_*L1, eq(lev)) h(1) robust nod.
```

The corresponding output is reported below.

```
Dynamic panel-data estimation, one-step system GMM
------------------------------------------------------------------------------
Group variable: id                              Number of obs      =       891
Time variable : year                            Number of groups   =       140
Number of instruments = 113                     Obs per group: min =         6
Wald chi2(12) =   4147.85                                      avg =      6.36
Prob > chi2   =     0.000                                      max =         8
------------------------------------------------------------------------------
             |               Robust
           n |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           n |
         L1. |   .8108394   .0579982    13.98   0.000     .6971649    .9245138
             |
           w |
```

[5] In order to run the `pca2` command more than once and to add GMM style IVs and PCA scores, the researcher is required to drop from the dataset the variables previously created by the command `pca2`: this can be done by typing, for example, `drop _BM* GMM*`.

```
         --. |  -.7945394   .0971517    -8.18   0.000    -.9849532   -.6041257
         L1. |     .55012    .151645     3.63   0.000     .2529012    .8473388
             |
           k |
         --. |   .4285055   .0763361     5.61   0.000     .2788895    .5781215
         L1. |  -.2802184   .0776689    -3.61   0.000    -.4324466   -.1279903
             |
    tauyear3 |   .0077488   .0200664     0.39   0.699    -.0315806    .0470781
    tauyear4 |    .020829   .0236973     0.88   0.379     -.025617    .0672749
    tauyear5 |  -.0002589   .0252166    -0.01   0.992    -.0496826    .0491648
    tauyear6 |  -.0271456     .02961    -0.92   0.359    -.0851801     .030889
    tauyear7 |   .0012306    .026954     0.05   0.964    -.0515983    .0540596
    tauyear8 |    .014436   .0254967     0.57   0.571    -.0355367    .0644087
    tauyear9 |   .0003278   .0307739     0.01   0.992     -.059988    .0606436
       _cons |   1.006162    .430149     2.34   0.019     .1630853    1.849238
------------------------------------------------------------------------------
Instruments for first differences equation
  Standard
    _GMML_k_1978L2 _GMML_k_1979L2 _GMML_k_1979L3 _GMML_k_1980L2 _GMML_k_1980L3

    (output omitted)

    D.(tauyear3 tauyear4 tauyear5 tauyear6 tauyear7 tauyear8 tauyear9)
Instruments for levels equation
  Standard
    _GMMD_k_1978L1 _GMMD_k_1979L1 _GMMD_k_1980L1 _GMMD_k_1981L1 _GMMD_k_1982L1

    (output omitted)

    tauyear3 tauyear4 tauyear5 tauyear6 tauyear7 tauyear8 tauyear9
    _cons
------------------------------------------------------------------------------
Arellano-Bond test for AR(1) in first differences: z =  -6.49  Pr > z =  0.000
Arellano-Bond test for AR(2) in first differences: z =  -0.08  Pr > z =  0.934
------------------------------------------------------------------------------
Sargan test of overid. restrictions: chi2(100)  = 113.34  Prob > chi2 =  0.171
  (Not robust, but not weakened by many instruments.)
Hansen test of overid. restrictions: chi2(100)  = 115.73  Prob > chi2 =  0.135
  (Robust, but weakened by many instruments.)
```

Similarly, we can get the SYS GMM estimates with the set of PCA scores from *PCIVR* (*_BM_varscoreDIF** and *_BM_avgscoreDIF*, _BM_varscoreLEV** and *_BM_avgscoreLEV**) as follows:

```
xtabond2 n l.n l(0/1).(w k) tauyear3-tauyear9, ///
iv(_BM_varscoreLEVn*, eq(diff) pass) iv(_BM_varscoreLEVw*, eq(diff) pass) ///
iv(_BM_varscoreLEVk*, eq(diff) pass) iv(_BM_varscoreDIFn*, eq(lev) pass) ///
iv(_BM_varscoreDIFw*, eq(lev) pass) iv(_BM_varscoreDIFk*, eq(lev) pass) ///
iv(tauyear3-tauyear9, eq(both)) h(1) robust nod.
```

```
Dynamic panel-data estimation, one-step system GMM
------------------------------------------------------------------------------
```

```
Group variable: id                              Number of obs      =       891
Time variable : year                            Number of groups   =       140
Number of instruments = 51                      Obs per group: min =         6
Wald chi2(12) =   5587.27                                      avg =      6.36
Prob > chi2   =     0.000                                      max =         8
------------------------------------------------------------------------------
             |               Robust
           n |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           n |
         L1. |   .9016193   .0477017    18.90   0.000     .8081257     .995113
             |
           w |
         --. |   -.742429   .1542546    -4.81   0.000    -1.044763   -.4400956
         L1. |   .4643432   .1950932     2.38   0.017     .0819675    .8467189
             |
           k |
         --. |     .53362    .096368     5.54   0.000     .3447423    .7224978
         L1. |  -.4411184   .1025934    -4.30   0.000    -.6421978    -.240039
             |
    tauyear3 |  -.0025501   .0226948    -0.11   0.911    -.0470312     .041931
    tauyear4 |   .0129266   .0272024     0.48   0.635    -.0403891    .0662423
    tauyear5 |   .0004112   .0272325     0.02   0.988    -.0529634    .0537858
    tauyear6 |  -.0197377   .0340792    -0.58   0.562    -.0865317    .0470564
    tauyear7 |   .0179079   .0346922     0.52   0.606    -.0500877    .0859034
    tauyear8 |   .0328657   .0278118     1.18   0.237    -.0216443    .0873758
    tauyear9 |      .0287   .0339306     0.85   0.398    -.0378027    .0952028
       _cons |   .9899051   .3951924     2.50   0.012     .2153422    1.764468
------------------------------------------------------------------------------
Instruments for first differences equation
  Standard
    _BM_varscoreLEVkN1 _BM_varscoreLEVkN2 _BM_varscoreLEVkN3

(output omitted)

    D.(tauyear3 tauyear4 tauyear5 tauyear6 tauyear7 tauyear8 tauyear9)
Instruments for levels equation
  Standard
    _BM_varscoreDIFkN1 _BM_varscoreDIFkN2 _BM_varscoreDIFkN3

(output omitted)

    tauyear3 tauyear4 tauyear5 tauyear6 tauyear7 tauyear8 tauyear9
    _cons
------------------------------------------------------------------------------
Arellano-Bond test for AR(1) in first differences: z =  -5.56  Pr > z =  0.000
Arellano-Bond test for AR(2) in first differences: z =  -0.27  Pr > z =  0.785
------------------------------------------------------------------------------
Sargan test of overid. restrictions: chi2(38)   =  57.54  Prob > chi2 =  0.022
  (Not robust, but not weakened by many instruments.)
Hansen test of overid. restrictions: chi2(38)   =  57.60  Prob > chi2 =  0.022
  (Robust, but weakened by many instruments.)
```

Even in this case, we observe a drop in the degrees of freedom of the

Hansen test that fall from 100 in the standard SYS GMM estimates to 38 when the scores relative to the principal components are extracted according to the variance criterion.

To further clarify the syntax and the options, we provide additional examples for the creation of instrumental variables and scores.

The syntax

```
pca2 n w k, nt(id year) lagsl(n, ll(2)) lagsl(w k,  ll(3))
```

creates PCA scores according to the variance criterion (90%) for each variable taken separately: the principal components are extracted both from the set of instruments in levels for n, which includes lags from $t-2$ up to the last lag available, and from the two sets of instruments in levels respectively for w and k that include lags from $t-3$ up to the last lag available.

The syntax

```
pca2 n w k, nt(id year) gmmd(2) lagsl(n w k, ll(2 3))
```

applies the PCA on the sets of instruments in first-differences from $t-2$ up to the last lag available for each variable taken separately and on the set of instruments in levels from $t-2$ to $t-3$ for each variable.

The syntax

```
pca2 n w k, nt(id year) lagsd(n w k, ll(2)) gmml(2) togvar togld var(80) avg
```

run the PCA on the set of instruments that includes the lags of interest both in levels and in first differences of all the variables taken together. The principal components are retained according to both the average and the variance (80%) criteria.

It is worth noticing that the syntax

```
pca2 n w k
```

pools all the observations and runs the PCA on the 3-columns matrix of n, w and k. It retains the principal components according to the variance criterion. The difference with respect to the syntax `pca n w k` is that the `pca2` also selects the principal components to be retained and computes the corresponding scores without the need of additional command lines.

4 `pca2` at work: an application to the estimation of a fiscal policy rule

In this section we illustrate more in detail the empirical implications and the operational advantages of the proposed procedure by applying it to the estimation of fiscal policy rules, as discussed in the paper by Golinelli and

Momigliano [2009] - GM henceforth[6]. In their work, the authors assess the robustness of the estimates of a fiscal policy rule on a panel of eleven Eurozone Countries over the post-Maastricht period (i.e. 1994-2008) by using alternative model specifications and exploiting data from different sources (European Commission, IMF, and OECD) and data vintages (latest available and real time data). Of main interest here, GM run a number of alternative regressions to estimate the parameters of the rule by SYS GMM and provide extensive motivations for their choice which comes out as the most appropriate in this framework, in line with well established indications in the literature[7]. However, we have stressed in previous sections that, when the cross-sectional dimension is smaller than the time dimension, there is the risk of getting biased estimates in case of a high number of over-identifying restrictions. Since the GM's dataset spans over N=11 and T=15, their analysis lacks of robustness checks with respect the number of orthogonality conditions exploited by the SYS GMM.

In this section we estimate the discretionary policy rule reported in GM [2009, p. 45]:

$$\Delta CAPB_{it} = \mu_i + \tau_t + \beta_1 GAP_{i,t-1} + \beta_2 CAPB_{i,t-1} + \beta_3 DEBT_{i,t-1} + \varepsilon_{it} \qquad (9)$$

where the dependent variable is the change in the cyclically adjusted primary borrowing on potential GDP measured with the latest available data (i.e. the best measure over time of the fiscal policy stance). The explanatory variables are the output gap (*GAP*), that accounts for the economic cycle, the cyclically adjusted primary borrowing (*CAPB*) and the debt (*DEBT*) as ratios on potential *GDP*, the latter two capturing the fiscal initial conditions. The explanatory variables are specified in $t-1$ and measured with real time data (i.e. the information available at the time when the fiscal policy is set). Finally, μ_i are Country fixed effects, τ_t are time fixed effects and ε_{it} are random policy shocks assumed to be i.i.d.

Table (1) reports estimates for equation (9) under alternative specifications of the over-identifying restrictions exploited by SYS GMM. We first estimate the model including all the available lags (col. 1); we then reduce the instrument set by lag-truncation (col. 2 and 3) and collapse (col. 4); finally, we exploit as new IVs the scores from *PCIVR* on the full set of instruments (col. 5), on the truncated set (col. 6) and on all lags of all the endogenous variables considered together (col. 7).

In particular, the estimates reported in the first column follow the same approach as in GM and are obtained by instrumenting all the explanatory variables with all the available lags as in GM (Table 3, column 5, "OECD-HP")[8].

[6]We are grateful to the authors for kindly providing us with their data

[7]The empirical analysis exploits the user-written Stata command `xtabond2`.

[8]These estimates do not perfectly match those in GM as here the *DEBT* is not considered as strictly exogenous.

These outcomes are in line with GM's ones: overall, the authors interpret their evidence as indicating that the fiscal initial conditions do affect policy choices, while the counter-cyclicality of fiscal policies comes out to be only slightly significant. It is worth noticing that in column 1 the number of over-identifying restrictions (152) is very close to the number of observations (165) and that the p-value of the Sargan test is almost 0.7: in the light of Sargan's caveats and of the discussion of the previous sections, we argue that the outcome of the test for over-identifying restrictions may be weakened by the high instrument count with respect to the number of observations.

In order to assess the robustness of GM findings, columns 2 to 7 of Table 1 report estimates where the number of over-identifying restrictions is reduced adopting alternative approaches.

More precisely, in column 2 the lag depth of the instruments is truncated to include only the lags for t-2 and t-3; in column 3 only the lags for t-2 are considered. Finally, the estimates in column 4 exploit a collapsed instrument matrix. Not surprisingly, the p-values of the Sargan test decrease with the instrument count, but this drop is also associated to substantial changes in the overall picture that emerges from the original GM estimates and from those in column 1 of Table 1. Quite strikingly, the two most frequently used approaches to reduce the instrument count contrast the GM findings in opposite directions. Indeed, the lag-truncation leads to estimates that strengthen the evidence of a-cyclicality of the fiscal policy, as the *GAP* parameter is not significant; the collapse of the instruments matrix gives not significant coefficient estimates for the *DEBT*, while the *GAP* parameter reaches a significance level of 10% and appears to some extent supportive of the counter-cyclicality of fiscal policies.

Such mixed evidence substantiates our concerns over the importance of introducing more compelling robustness checks in the cases when the instrument count is high and needs to be reduced.

To investigate the issue further, the last three columns in Table 1 report results for the estimating equation when the IVs count is reduced using the *PCIVR* strategy, as implemented through the `pca2` command. In column 5 the SYS GMM estimator exploits as IVs the PCA scores relative to the principal components of the matrix that includes all the available lags, retained according to the average criterion; this strategy is less parsimonious than the collapsing in terms of number of moment restrictions and it provides results that are in line with those obtained on the whole instrument set. In column 6 the principal components to be retained according to the average criterion for the computation of the scores are extracted from the instruments matrix that includes only the lags relative to $t-2$ and $t-3$. It is also interesting to remark that the number of degrees of freedom of the Sargan test is larger than that obtained from the collapsing. In order to reduce the instrument count to a number in line with that of the collapse, in the estimates of column 7

the principal components are extracted from both the instrument matrix that includes all the lags in levels of all the variables taken together and from that with all the IVs in first-differences[9].

Overall, the empirical exercise performed in this section conveys important empirical indications. First, we see that the estimates on the sets of instruments obtained through the `pca2` command are in line with the findings of columns (1)-(3), providing at the same time a lower instrument count and a higher reliability of the Sargan test which is consistently characterised by a lower p-value. The results provided here corroborate the idea that the *PCIVR*, being a purely statistical way to tackle the issue of the excess of counts, has the advantage of doing that without imposing heavy (and somewhat arbitrary) restrictions on the data structure. This feature emerges in particular from column 7, whose outcome closely mirrors the one in column 1 but is now obtained with a number of over-identifying restrictions that is only one-third compared to the former one.

Finally, with respect to the policy implications of the GM study, we have shown that estimates based on collapsed instruments would have changed the view over the determinants of the policy rules, as the stock of debt is found as non significant in contrast with the significant coefficient across all the others specifications. Thanks to the newly implemented `pca2` procedure, we have been able to show that this shift is not directly driven by the reduction in the number of IVs (also carried out by lag-truncation and *PCIVR*) but it is rather due to the restrictions imposed on the instrument matrix.

[9]This is done by specifying the option `togvar` for the `pca2` command. The data and the `.do` file with the commands to replicate Table 1 are provided as complementary material.

Table 1: Estimation of a fiscal policy rule

Dependent variable: $\Delta CAPB$								
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
Variable		all lags	lags 2-3	lag 2	collapse	PCIVR all	PCIVR col	PCIVR tog
L.CAPB	*coeff*	-0.317	-0.306	-0.303	-0.382	-0.318	-0.306	-0.334
	sd	0.058	0.06	0.064	0.089	0.062	0.067	0.08
	t	-5.43	-5.12	-4.72	-4.3	-5.13	-4.6	-4.15
L.DEBT	*coeff*	0.017	0.016	0.014	0.017	0.014	0.014	0.021
	sd	0.004	0.005	0.005	0.017	0.005	0.006	0.009
	t	3.91	3.41	2.76	0.96	2.73	2.58	2.38
L.GAP	*coeff*	0.146	0.131	0.128	0.238	0.16	0.097	0.21
	sd	0.096	0.099	0.107	0.137	0.1	0.114	0.146
	t	1.52	1.32	1.2	1.74	1.6	0.85	1.44
Constant	*coeff*	-0.757	-0.635	-0.532	-0.703	-0.496	-0.556	-1.03
	sd	0.454	0.46	0.485	1.221	0.478	0.512	0.694
	t	-1.67	-1.38	-1.1	-0.58	-1.04	-1.09	-1.48
NxT		165	165	165	165	165	165	165
N		11	11	11	11	11	11	11
T		15	15	15	15	15	15	15
Sargan test:								
degr. of fr.		152	132	87	48	117	83	46
P value		0.6928	0.6246	0.1612	0.2276	0.2691	0.2048	0.1451

References

[1] **Altonji, J.G. and Segal, L.M. (1996)** "Small-Sample Bias in GMM Estimation of Covariance Structures", *Journal of Business and Economic Statistics*, Vol. 14, pp. 353-366.

[2] **Alvarez, J. and Arellano, M. (2003)** "The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators", *Econometrica*, Vol. 71(4), pp. 1121-1159.

[3] **Amemiya, T. (1966)** "On the use of principal components of independent variables in two-stage least-squares estimation", *International Economic Review*, Vol. 7, pp. 283-303.

[4] **Andersen, T.G. and Sorensen, B.E (1996)** "GMM Estimation of a Stochastic Volatility Model: A Monte Carlo Study", *Journal of Business and Economic Statistics*, American Statistical Association, Vol. 14(3), pp. 328-52.

[5] **Arellano, M. and S.R. Bond (1991)** "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations", *Review of Economic Studies*, Vol. 58, pp. 277-297.

[6] **Arellano, M., O. Bover (1995)** "Another look at the instrumental variables estimation of error-components models", *Journal of Econometrics*, vol. 68, pp. 29-51.

[7] **Bai, J. and S. Ng (2010)** "Instrumental Variable Estimation In A Data Rich Environment", *Econometric Theory*, Vol. 26, pp. 1577-1606.

[8] **Bekker, P.A. (1994)** "Alternative approximations to the distributions of instrumental variable estimators", *Econometrica*, Vol. 62, pp. 657-681.

[9] **Blundell, R.W. and S.R. Bond (1998)** "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models", *Journal of Econometrics*, Vol. 87, pp. 115-143.

[10] **Bound, J., Jaeger, D.A. and Baker, R.M. (1995)** "Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak", *Journal of the American Statistica Association*, vol. 90, pp. 443-450.

[11] **Bowsher, C. G. (2002)** "On testing overidentifying restrictions in dynamic panel data models", *Economics Letters*, vol. 77, pp. 211-220.

[12] **Dagenais, M.G. and Dagenais, D.L. (1997)** "Higher moment estimator for linear regression models with errors in the variables", *Journal of Econometrics*, Vol. 76, pp. 193-221.

[13] **Doran H.E. and Schmidt P. (2006)** "GMM estimators with improved finite sample properties using principla components of the weighting matrix, with an application to the dynamic panel data model", *Journal of econometrics*, Vol. 133, pp. 387-409.

[14] **Golinelli, R. and S. Momigliano (2009)** "The cyclical reaction of fiscal policies in the Euro area: the role of modeling choices and data vintages", *Fiscal Studies*, vol. 30, n. 1.

[15] **Groen, J.J.J. and G. Kapetanios (2009)** "Parsimonious estimation with many instruments", *Federal Reserve Bank of New York*, Staff Report n. 386.

[16] **Han C. and P.C.B. Phillips (2006)** "GMM with many moment conditions", *Econometrica*, Vol. 74, pp. 147-192.

[17] **Kapetanios, G. and M. Marcellino (2010)** "Factor-GMM estimation with large sets of possibly weak instruments manuscript", *Computational Statistics and Data Analysis*, Vol. 54, pp. 2655-2675.

[18] **Kiviet, J.F. (1995)** "On bias, inconsistency, and efficiency of various estimators in dynamic panel data models", *Journal of Econometrics*, Vol. 68(1), pp. 53-78

[19] **Kloek, T., and L.B.M. Mennes (1960)** "Simultaneous equations estimation based on principal components of predetermined variables", *Econometrica*, Vol. 28, pp. 45-61.

[20] **Mehrhoff, J., (2009)** "A solution to the problem of too many instruments in dynamic panel data GMM", *Discussion paper n. 1/2009*, Deutsche Bundesbank.

[21] **Newey, W.K. and Smith, R.J. (2004)** "Higher-order properties of GMM and generalized empirical likelihood estimators", *Econometrica*, Vol. 72, pp.219-255.

[22] **Roodman, D. (2009a)** "A Note on the theme of too many instruments", *Oxford Bulletin of Economics and Statistics*, Vol. 71, pp. 135-158.

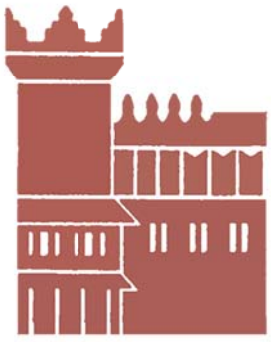
[23] **Roodman, D. (2009b)** "How to Do xtabond2: An Introduction to "Difference" and "System" GMM in Stata", *Stata Journal*, Vol. 9, pp. 86-136.

[24] **Sargan, J.D. (1958)** "The estimation of economic relationships using instrumental variables", *Econometrica*, Vol. 26, pp. 393-415.

[25] **Staiger, D. and Stock, J.H., (1997)** "Instrumental Variables Regression with Weak Instruments", *Econometrica*, Vol. 65, pp. 557-586.

[26] **Stock, J.H. and M.W. Watson, (2002)** "Forecasting using principal components from a large number of predictors", *Journal of the American Statistical Association*, Vol. 97, pp. 1167-1179.

[27] **Ziliak J. P. (1997)** "Efficient estimation with panel data when instruments are predetermined: an empirical comparison of moment-condition estimators", *Journal of Business and Economic Statistics*, Vol. 15, pp. 419-431.



Alma Mater Studiorum - Università di Bologna
DEPARTMENT OF ECONOMICS

Strada Maggiore 45
40125 Bologna - Italy
Tel. +39 051 2092604
Fax +39 051 2092664
http://www.dse.unibo.it